In [14]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
from zipfile import ZipFile
file_name = "DecisionTreeAssignmentProblem-210629-152345.zip"
```

In [3]:
```
file_name
```

Out[3]:
```
'DecisionTreeAssignmentProblem-210629-152345.zip'
```

In [9]:
```
file_name
with ZipFile(file_name, 'r') as zip:
```

```
  Input In [9]
    with ZipFile(file_name, 'r') as zip:
                                         ^
IndentationError: expected an indented block
```

In [7]:
```
file_name2 = pd.read_zip(r'DecisionTreeAssignmentProblem-210629-152345.zip')
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Input In [7], in <cell line: 1>()
----> 1 file_name2 = pd.read_zip(r'DecisionTreeAssignmentProblem-210629-152345.
zip')

File ~\anaconda3\lib\site-packages\pandas\__init__.py:261, in __getattr__(name)
    257     from pandas.core.arrays.sparse import SparseArray as _SparseArray
    259     return _SparseArray
--> 261 raise AttributeError(f"module 'pandas' has no attribute '{name}'")

AttributeError: module 'pandas' has no attribute 'read_zip'
```

In [12]:
```
from zipfile import ZipFile

# specifying the zip file name
file_name = "DecisionTreeAssignmentProblem-210629-152345.zip"

# opening the zip file in READ mode
with ZipFile(file_name, 'r') as zip:
    # printing all the contents of the zip file
    zip.printdir()

    # extracting all the files
    print('Extracting all the files now...')
    zip.extractall()
    print('Done!')
```

```
File Name                                             Modified             Size
Decision Tree Assignment Problem.ipynb        2021-05-19 04:59:16        20484
Extracting all the files now...
Done!
```

In [13]:
```
file_name
```

Out[13]:
```
'DecisionTreeAssignmentProblem-210629-152345.zip'
```

In [15]:
```
data = pd.read_csv(filepath_or_buffer='https://raw.githubusercontent.com/insaid20
```

In [16]:
```
data
```

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	

891 rows × 12 columns

In [17]:
```
print('shape of the data is : ',data.shape)
```

```
shape of the data is :  (891, 12)
```

In [19]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [21]:
```
data.describe()
```

Out[21]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [ ]:
```
#"observations from the above function"
#there are total of 891 passengers
#less than 50 % of passengers survived the mishap
#more 75 % passengers travelling in 3rd class
#25 % passengers travelling in 2rd class
#max age is 80, min age is 0.42, and averege age is 29
#50 % passengers have no siblings/spouce
#max siblings/spouce is 8
#more than 75 % have no parents/children
#max fare is 512.32, min fare is 0.00 and averege fare is 32.20
```

In [22]:
```
data.head(10)
```

Out[22]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Na
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	Na
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C12
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	Na
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	Na
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E4
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	Na
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	Na
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	Na

In [33]:
```
!pip install -q datascience
!pip install -q pandas-profiling
!pip install -q yellowbrick

!pip install -q --upgrade pandas-profiling
!pip install -q --upgrade yellowbrick

import pandas as pd
from pandas_profiling import ProfileReport
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
pd.set_option('mode.chained_assignment', None)
```

```
---------------------------------------------------------------------------
ImportError                               Traceback (most recent call last)
Input In [33], in <cell line: 9>()
      6 get_ipython().system('pip install -q --upgrade yellowbrick')
      8 import pandas as pd
----> 9 from pandas_profiling import ProfileReport
     10 pd.set_option('display.max_columns', None)
     11 pd.set_option('display.max_rows', None)

File ~\anaconda3\lib\site-packages\pandas_profiling\__init__.py:6, in <module>
      1 """Main module of pandas-profiling.
      2
      3 .. include:: ../../README.md
      4 """
----> 6 from pandas_profiling.controller import pandas_decorator
      7 from pandas_profiling.profile_report import ProfileReport
      8 from pandas_profiling.version import __version__

File ~\anaconda3\lib\site-packages\pandas_profiling\controller\pandas_decorato
r.py:4, in <module>
      1 """This file add the decorator on the DataFrame object."""
      2 from pandas import DataFrame
----> 4 from pandas_profiling.profile_report import ProfileReport
      7 def profile_report(df: DataFrame, **kwargs) -> ProfileReport:
      8     """Profile a DataFrame.
      9
     10     Args:
   (...)
     15         A ProfileReport of the DataFrame.
     16     """

File ~\anaconda3\lib\site-packages\pandas_profiling\profile_report.py:27, in <m
odule>
     25 from pandas_profiling.report.presentation.core import Root
     26 from pandas_profiling.report.presentation.core.renderable import Render
able
---> 27 from pandas_profiling.report.presentation.flavours.html.templates impor
t (
     28     create_html_assets,
     29 )
     30 from pandas_profiling.serialize_report import SerializeReport
     31 from pandas_profiling.utils.dataframe import hash_dataframe

File ~\anaconda3\lib\site-packages\pandas_profiling\report\presentation\flavour
```

```
s\html\__init__.py:1, in <module>
----> 1 from pandas_profiling.report.presentation.flavours.html.alerts import H
TMLAlerts
      2 from pandas_profiling.report.presentation.flavours.html.collapse import
HTMLCollapse
      3 from pandas_profiling.report.presentation.flavours.html.container impor
t HTMLContainer

File ~\anaconda3\lib\site-packages\pandas_profiling\report\presentation\flavour
s\html\alerts.py:2, in <module>
      1 from pandas_profiling.report.presentation.core.alerts import Alerts
----> 2 from pandas_profiling.report.presentation.flavours.html import template
s
      5 class HTMLAlerts(Alerts):
      6     def render(self) -> str:

File ~\anaconda3\lib\site-packages\pandas_profiling\report\presentation\flavour
s\html\templates.py:5, in <module>
      2 import shutil
      3 from pathlib import Path
----> 5 import jinja2
      7 from pandas_profiling.config import Settings, Theme
      8 from pandas_profiling.report.formatters import fmt, fmt_badge, fmt_nume
ric, fmt_percent

File ~\anaconda3\lib\site-packages\jinja2\__init__.py:12, in <module>
     10 from .bccache import FileSystemBytecodeCache
     11 from .bccache import MemcachedBytecodeCache
---> 12 from .environment import Environment
     13 from .environment import Template
     14 from .exceptions import TemplateAssertionError

File ~\anaconda3\lib\site-packages\jinja2\environment.py:25, in <module>
     23 from .compiler import CodeGenerator
     24 from .compiler import generate
---> 25 from .defaults import BLOCK_END_STRING
     26 from .defaults import BLOCK_START_STRING
     27 from .defaults import COMMENT_END_STRING

File ~\anaconda3\lib\site-packages\jinja2\defaults.py:3, in <module>
      1 # -*- coding: utf-8 -*-
      2 from ._compat import range_type
----> 3 from .filters import FILTERS as DEFAULT_FILTERS  # noqa: F401
      4 from .tests import TESTS as DEFAULT_TESTS  # noqa: F401
      5 from .utils import Cycler

File ~\anaconda3\lib\site-packages\jinja2\filters.py:13, in <module>
     11 from markupsafe import escape
     12 from markupsafe import Markup
---> 13 from markupsafe import soft_unicode
     15 from ._compat import abc
     16 from ._compat import imap

ImportError: cannot import name 'soft_unicode' from 'markupsafe' (C:\Users\user
\anaconda3\lib\site-packages\markupsafe\__init__.py)
```

In [34]:
```
profile = ProfileReport(df=data)
profile.to_file(output_file='Pre Profiling Report.html')
print('Accomplished!')
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [34], in <cell line: 1>()
----> 1 profile = ProfileReport(df=data)
      2 profile.to_file(output_file='Pre Profiling Report.html')
      3 print('Accomplished!')

NameError: name 'ProfileReport' is not defined
```

In [35]:
```
data.columns
```

Out[35]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [52]:
```
#inserting missing values
data['Embarked']=data['Embarked'].fillna(value=data['Embarked'].mode()[0])
data['Age'].fillna(value=data['Age'].median(),inplace = True)
data.drop(labels='Cabin',axis=1,inplace=True)
```

In [53]:
```
data.head()
```

Out[53]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Emb
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	

In [54]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          891 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Embarked     891 non-null    object 
dtypes: float64(2), int64(5), object(4)
memory usage: 76.7+ KB
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: